Statistical Evaluation and Review

Date:	August 23, 2001
Type:	BLA
FDA number:	99-1492 (STN 103951/0.0)
Product/Application:	NESP (ARANESP), novel erythropoietin stimulating protein.
Indication:	Treatment of anemia in ESRD patients.
Sponsor:	Amgen, Inc.
Title of Document:	BLA Submission dated Dec 23, 1999.
From:	Ghanshyam Gupta, Ph.D., HFM-219 Ghanshyam Gupta 8-23-01
Through:	P. A. Lachenbruch, Ph.D., Director, Division of Biostatistics, HFM-215 P Lachenbruch
To:	Ellis Unger, M. D., HFM-576
	Lori Tull, HFM-585
	Chron,
	DCC, HFM-99
	S. Ellenberg, HFM-210 sr yarm

Background:

NESP is a novel erythropoiesis stimulating protein produced in Chinese hamster ovary (CHO) cells by recombinant DNA technology. It is a 165-amino acid protein containing 5 N-linked oligosaccharide chains, whereas endogenous erythropoietin (EPO) or recombinant human erythropoietin (rHuEPO) contains only 3. It stimulates erythropoiesis by the same mechanism as endogenous EPO or rHuEPO. The pharmacokinetic studies have shown that it has a 2 to 3-fold longer half-life than rHuEPO and consequently a greater in vivo activity when administered by either the intravenous or subcutaneous route.

Submission:

The submission consists of 13 clinical studies. There are two pivotal studies in the submission. One of these two studies was conducted in the US (NESP 980117) and the other study was an open label European study (NESP 970200). This review is mostly

based on these two studies. Both of the studies were designed to show non-inferiority of NESP in comparison to rHuEPO.

Study NESP 980117

This was a multi-enter, randomized, double-blind, non-inferiority study of NESP administered intravenously (IV) once weekly versus rHuEPO administered IV three times weekly in chronic renal failure (RCF) patients with anemia. After randomization, patients were maintained within a hemoglobin target range of -1.0 g/dL to +1.5 g/dL from their baseline hemoglobin value. A period of 20 weeks after the first dose of study drug was used for dose-titration and stabilization of hemoglobin. Efficacy endpoints were assessed during the evaluation period (weeks 21 to 28).

The primary endpoint of the study was a change in hemoglobin from baseline through the evaluation period. The statistical hypothesis was that the lower limit of the 2-sided 95% confidence interval for the difference between the NEST and EPO groups in mean change in hemoglobin was above -1.0 g/dL. The primary analysis is based on the per-protocol population and other analyses were considered secondary. Some of the important secondary endpoints of the study were:

- Instability of hemoglobin during the evaluation period.
- Percentage of hemoglobin values within the target range (-1.0 to +1.5) from baseline and 9.0 to 13.0 g/dL during the evaluation period.
- Percentage of hemoglobin values within the therapeutic range of 9.0 to 13.0 g/dL during the evaluation period.
- Dose of the study drug during the evaluation period.

Study Results

The study randomized 507 patients, 169 of whom were assigned to NESP and 338 were assigned to EPO. The randomization was supposed to assign patients in 2:1 ratio with twice the number of patients to the NESP group. However, patient randomization was inadvertently reversed in the 1:2 ratio. The sponsor informed the FDA of the mistake.

The FDA informed the sponsor to continue the study without any amendment. The study was conducted in 35 centers in the US and 5 centers in Canada.

Baseline Comparisons

At baseline the two treatment groups were found to be comparable with respect to sex, race, age, and baseline hemoglobin levels. The mean age of patients was 57.8 years in the EPO group verses 58.0 in the NESP group (p=0.93). There were 56.5% males in the EPO group verses 55.6% males in the NESP group. There were slightly higher percentages of patients with age ≥ 65 years in the EPO group then the NESP group (40% versus 31%, p>0.05). Other baseline details are provided below in Table 1.

Table 1. Baseline Characteristics

Characteristic	EPO	NESP	P-value
SEX: Female	147 (43.5%)	75 (44.4%)	0.85
Male	191 (56.5%)	94 (55.6%)	
Age <65 yrs	204 (60.4%)	117 (69.2%)	0.0506
>=65 yrs	134 (39.6%	52 (30.8%)	
Age <75 yrs	292 (86.4%)	145 (85.8%)	0.86
>=75 yrs	46 (13.6%)	24 (14.2%)	
Race: Caucasian	144 (42.6%)	68 (40.2%)	0.84
Black	129 (38.2%)	69 (40.8%)	
Other	65 (19.2%)	32 (18.9%)	
Hemoglobin <=10.3	41 (12.1%)	18 (10.7%)	0.29
10.3 to <=11.5	169 (50.0%)	54 (57.4%)	
>11.5	128 (37.9%)	54 (32.0%)	
Country: Canada	48 (14.2%)	24 (14.2%)	1.00
USA	290 (85.8%)	145 (85.8%)	

The Primary Efficacy Endpoint Analysis

According to the protocol, the primary comparison involved analysis of covariance of change from baseline in hemoglobin levels during the maintenance period of 21 to 28th week with baseline hemoglobin levels and center as covariables. The sponsor provided

three data sets (1) OBSERVED: actual data, (2) LVCF: last value carried forward if missed, and (3) SLVCF: some data use substitute, the rest use LVCF. The sponsor also supplied an imputed data set. This reviewer did not analyze this data set. We analyzed the primary endpoint of change in hemoglobin using all the other three data sets. In all three analyses, treatment groups were not significantly different. Baseline hemoglobin levels as expected were highly significant. Center effect was also found significant in all three data sets. The adjusted mean in change in hemoglobin levels for the three data sets is provided below in Table 2.

Table 2. Adjusted Mean Change in Hemoglobin Levels (g/dL) during the Evaluation Period.

DATASET	EPO Mean ± SE	NESP Mean ± SE	Difference (NESP-EPO)	95% CI
OBSERVED	-0.129 ± 0.065 n=293	-0.043 ± 0.087 n=150	0.085	-0.120, 0.290
LVCF	-0.137 ± 0.065 n=334	-0.049 ± 0.089 n=169	0.088	-0.118, 0.294
SLVCF	-0.194 ± 0.065 n=334	-0.137 ± 0.088 n=169	0.057	-0.149, 0.263

All three sets of analyses showed that the lower limit of the 95% confidence interval of the difference between the adjusted mean change in hemoglobin levels was well above the prospectively defined -1.0 g/dL level. The adjusted mean change from baseline was larger in the EPO group by about 0.08 g/dL in the 'observed' and 'lvcf' data sets. In addition to the expected significant effect of the baseline hemoglobin levels, the center as well as center by treatment effects was also found significant. However, due to large number of centers and varied number of patients per center, it was difficult to make any judgement from the significant interaction effect. The sponsor pooled several small centers to homogenize variation due to center effect prior to conducting analysis of covariance analyses.

In addition to the prospective analysis of covariance, we analyzed the three datasets without any adjustment. These results are given below in Table 3.

Table 3. Mean Change in Hemoglobin Levels (g/dL) during the Evaluation Period.

DATASET	EPO Mean ± SE	NESP Mean ± SE	Difference (NESP-EPO)	95% CI
OBSERVED	-0.125 ± 0.063 n=293	-0.013 ± 0.091 n=150	0.113	-0.328, 0.103
LVCF	-0.141 ± 0.062 n=334	-0.036 ± 0.093 n=169	0.105	-0.320, 0.110
SLVCF	-0.204 ± 0.063 n=334	-0.132 ± 0.093 n=169	0.071	-0.288, 0.145

Except for the small change in the difference estimate, this analysis also provides results very similar to the one obtained by analysis of covariance.

Mean Hemoglobin Levels during the Evaluation Period

We analyzed mean hemoglobin levels also during the evaluation period of weeks 21 to 28. As before, we analyzed all three data sets. The results are provided below in Table 4.

Table 4. Mean Hemoglobin Levels (g/dL) during the Evaluation Period.

DATASET	EPO Mean ± SE	NESP Mean ± SE	Difference (NESP-EPO)	95% CI
OBSERVED	11.121 ± 0.065 n=293	11.124 ± 0.094 n=150	0.003	-0.227, 0.220
LVCF	11.089 ± 0.064 n=334	11.113 ± 0.096 n=169	0.024	-0.247, 0.200
SLVCF	11.027 ± 0.065 n=334	11.017 ± 0.095 n=169	0.011	-0.212, 0.234

The mean hemoglobin levels during the evaluation period were very much alike in both groups. The 95% confidence interval of the difference was also seen within a narrow range.

Subset Analyses of the Change in Hemoglobin levels from the Baseline

Subset analyses by sex, and age (< 65, ≥65 years) of the change in hemoglobin levels from baseline as well as of the mean hemoglobin levels during the evaluation period were also conducted. In addition, interaction effects with treatment were also evaluated for each subset variable. No significant interaction effects were observed. The mean and standard errors along with 95% CI for the observed and LVCF data sets are provided below in Table 5 and 6. The third data set SLVCF being fairly similar to LVCF was not analyzed.

Table 5. Subset Analysis of the Change in Hemoglobin Levels during the Evaluation Period

DATASET	SUBSET VARIABLE		EPO	NESP	DIFFERENCE (NESP-EPO)	95% CI
OBSERVED	SEX	MALE	N=172 -0.13 ± 0.08	N=81 -0.08 ± 0.13	0.05	-0.35, 0.25
		FEMALE	N=121 -0.12 ± 0.09	N=69 **0.06 ± 0.13**	0.18	-0.50, 0.13
	AGE	< 65 yrs	N=184 -0.13 ± 0.09	N=102 0.01 ± .011	0.14	-0.42, 0.14
		≥ 65 yrs	N=109 -0.12 ± 0.09	N=48 -0.06 ± 0.16	0.05	-0.40, 0.29
LVCF	SEX	MALE	N=190 -0.14 ± 0.08	N=94 -0.05 ± 0.13	0.09	-0.38, 0.19
		FEMALE	N=144 -0.14 ± 0.10	N=75 -0.02 ± 0.13	0.12	-0.45, 0.20
	AGE	< 65 yrs	N=202 -0.11 ± 0.09	N=117 -0.04 ± 0.11	0.07	-0.35, 0.21
		≥ 65 yrs	N=132 -0.18 ± 0.09	N=52 -0.02 ± 0.16	0.16	-0.50, 0.17

As seen in the table above, the decrease in hemoglobin from baseline was consistent within all subgroups except for females in the observed group where there was a slight increase in the mean change in hemoglobin. Hemoglobin data during the evaluation

period were also analyzed for the two data sets. The results are provided below in Table 6.

Table 6. Subset Analysis of the Hemoglobin Levels during the Evaluation Period

DATASET	SUBSET VARIABLE		EPO	NESP	NESP-EPO (95% CI)
OBSERVED	SEX	MALE	N=172 11.19 ± 0.09	N=81 11.13 ± 0.13	0.06 (-0.24, 0.37)
		FEMALE	N=121 11.02 ± 0.10	N=69 11.12 ± 0.13	0.10 (-0.43, 0.23)
	AGE	< 65 yrs	N=184 11.14 ± 0.08	N=102 11.15 ± .012	0.01 (-0.29, 0.27)
		≥ 65 yrs	N=109 11.08 ± 0.10	N=48 11.07 ± 0.16	0.02 (-0.35, 0.39)
LVCF	SEX	MALE	N=190 11.16 ± 0.08	N=94 11.15 ± 0.14	0.01 (-0.29, 0.31)
		FEMALE	N=144 10.99 ± 0.10	N=75 11.07 ± 0.13	0.07 (-0.40, 0.26)
	AGE	< 65 yrs	N=202 11.15 ± 0.09	N=117 11.10 ± 0.12	0.04 (-0.24, 0.33)
		≥ 65 yrs	N=132 11.00 ± 0.10	N=52 11.13 ± 0.16	0.13 (-0.49, 0.23)

Secondary Endpoints

Instability of the hemoglobin levels for patients during the evaluation period was the first secondary endpoint. We analyzed the standard deviations of the hemoglobin levels during the evaluation period using Wilcoxon Rank Sum test. For the observed data set, the difference between the two treatment groups reached almost significance (p=0.0625). However, with LOCF, instability as measured by standard deviation was increased slightly in the EPO group and the analysis showed that the two treatment groups were significantly different. The NESP group was observed to be slightly more variable.

Table 7. Analysis of SD of HgB during the Evaluation Period

DATASET	EPO		NESP		p-value*
	N	MEDIAN (RANGE)	N	MEDIAN (RANGE)	
OBSERVED	334	0.406 0 - 3.25	169	0.457 0 - 4.50	0.0625
LOCF	291	0.449 0.13 - 3.25	148	0.485 0.14 - 4.50	0.0426

* Wilcoxon rank-sum test

Percentages of hemoglobin values within the target range (-1.0 to +1.5) from baseline were another secondary variable. Hundred percent of the patients observed achieved this target within each of the treatment group. Percentage of patients with hemoglobin values within the therapeutic range of 9.0 and 13.0 g/dL was also a secondary efficacy variable. Percentage of patients who had their hemoglobin levels within the target range was about 90% in both the OBSERVED and LOCF datasets.

Table 8. Mean Hemoglobin Levels in the Target Range (9.0 to 13.0).

DATASET	EPO		NESP		p-value
	N	WITHIN RANGE	N	WITHIN RANGE	
OBSERVED	334	303 (90.7%)	169	150 (88.8%)	0.49
LOCF	293	269 (91.8%)	150	137 (91.3%)	0.86

However, the percentage of evaluations per patient which were within the therapeutic range during the evaluation period ranged from zero to 100 percent. In the observed dataset, there were 10 percent of patients with 62.5% compliance in the EPO group and 50% compliance in the NESP group. Similar results were observed with the LVCF dataset.

Treatment dose in the evaluation period in the EPO group varied from 134 to 26250 with mean dose of 4937 (SE=380) while in the NESP group it varied from 1.50 to 103 with a mean of 19.18 (SE=1.74).

Study NESP 970200

This was a European, open label, randomized, phase III non-inferiority study comparing EPO and NESP for the treatment of anemia in CRF patients receiving HD or PD. Patients were randomized to NESP and EPO in a 2:1 ratio. A period of 1-24 weeks after the first dose of study drug was used for dose titration and stabilization of hemoglobin. Efficacy endpoints were assessed during the evaluation period from week 25 to 32. There was an additional 20-week maintenance period for further safety comparisons.

The study enrolled 522 patients with 347 patients randomized to NESP and 175 patients randomized to EPO group. Thirty-one centers participated in the study. Of the 522 patients, 519 (99%) received study drug.

Demographic Comparisons

There were 288 (55%) males in the study. Caucasian constituted a large (92%) of the patients. Of the nine countries participating in the study, a large majority (61%) of the patients belonged to the four countries (UK, Australia, Germany, and France). The mean average of patients was about 60 years. The two treatments groups were not found to differ significantly with regard to any of these characteristics. The results of the comparisons are provided below in Table 9.

Table 9. Baseline Characteristics

Characteristic	EPO	NESP	P-value
SEX: Female	75 (42.9%)	159 (45.8%)	0.52
Male	100 (57.1%)	188 (54.1%)	
Age <65 yrs	97 (55.4%)	192 (55.3%)	0.98
>=65 yrs	78 (44.6%	155 (44.7%)	
Race: Caucasian	165 (94.3%)	316 (91.1%)	0.82
Other	10 (5.7%)	21 (8.9%)	
Hemoglobin* <=10.3	86 (49.1%)	175 (50.4%)	0.78
>10.3	89(50.9%)	172 (49.6%)	

* There were no patients with more than Hgb>11.5 g/dL.

Efficacy Results

The primary endpoint of the study was the same as in the US study NESP980117 with similar non-inferiority criterion. The mean difference in the change in hemoglobin from baseline during the evaluation period of weeks 25 to 32 for the observed patients was 0.06 with 95% confidence interval of (-0.25, 0.13). The lower limit of the 95% CI was well above the -0.5 g/dL prospectively defined limit for non-inferiority. Mean difference in the hemoglobin level was also well above -0.5 g/dL. These results are provided below in Table 10.

Table 10. Mean Change in Hemoglobin Levels (g/dL) during the Evaluation Period.

Efficacy Variable	EPO Mean ± SE	NESP Mean ± SE	Difference (NESP-EPO)	95% CI
Hemoglobin Level	N=158 10.95 ± 0.09	N=302 11.03 ± 0.06	0.07	-0.28, 0.13
Change from Baseline	N=158 0.07 ± 0.08	N=302 0.007 ± 0.06	0.06	-0.25, 0.13

Similar results were obtained when the data were analyzed by sex. The means and standard errors along with 95% confidence intervals of the hemoglobin and change in hemoglobin from baseline by are provided below in Table 11.

Table 11. Hemoglobin Levels (g/dL) and Change in Hgb from Baseline by Sex during the Evaluation Period.

Efficacy Variable	**Subgroup**	**EPO Mean ± SE**	**NESP Mean ± SE**	**Difference 95% CI**
Hemoglobin Level	Male	N=89 11.13 ± 0.11	N=162 11.11 ± 0.08	0.14 -0.26, 0.29
	Female	N=69 10.73 ± 0.13	N=140 10.93 ± 0.09	0.20 -0.51, 0.11
Change from Baseline	Male	N=89 0.05 ± 0.11	N=162 0.02 ± 0.07	0.13 -0.22, 0.28
	Female	N=69 -0.23 ± 0.11	N=140 -0.04 ± 0.09	0.19 -0.48, 0.10

In addition to the mean change in hemoglobin levels from the baseline, the variability of the hemoglobin levels during the evaluation period was also of interest. We analyzed the standard deviations of hemoglobin levels during the evaluation period for comparison of variability between the two groups. The median standard deviation was 0.38 for the EPO group versus 0.42 for the NESP group. The two groups were not found to be statistically significantly different with regard to standard deviation. Similar results were obtained for the standard deviation of the change in hemoglobin from baseline.

Table 12. Analysis of Standard Deviation of Hemoglobin Levels (g/dL) during the Evaluation Period (week 25-32).

Efficacy Variable	EPO Median (Range)	NESP Median (Range)	p-value*
Hemoglobin	N=158 0.38 (0.09, 1.55)	N=302 0.42 (0.10, 1.58)	0.1395
Change from Baseline	N=158 0.38 (0.09, 1.55)	N=302 0.42 (0.10, 1.58)	0.1386

* Wilcoxon Rank Sum Test

Safety

Analysis of adverse experiences of the two studies showed very similar profile. The top 10 adverse experiences from both the studies are provided below in Table 11. In the US study, nausea was the most common adverse experience while hypotension occurred most frequently in the European study. These followed by hypotension and upper respiratory infection in the European study, and by myalgia and hypertension in the US study. Overall, the adverse experience incidences appear slightly higher in the European study. However, the incidence rates of the two should be interpreted with the knowledge that the European study was an open label study while the US/Canada study was a double blind study. The frequently occurring adverse experiences did not appear to differ between the two treatment groups within each study.

Table 13. Top ten adverse experiences by study and treatment

Study: 980117			Study 970200		
AE	**EPO**	**NESP**	**AE**	**EPO**	**NESP**
Nausea	0.28	0.29	Hypotension	0.38	0.39
Hypertension	0.24	0.28	Myalgia	0.36	0.33
Upper Resp Inf	0.27	0.27	Hypertension	0.27	0.30
Dyspnea	0.20	0.26	Diarrhea	0.23	0.26
Myalgia	0.22	0.21	Headache	0.20	0.21
Pain Chest	0.15	0.21	Upper Resp Inf	0.21	0.20
Vomiting	0.21	0.20	Vomiting	0.25	0.19
Edema Perphera	0.19	0.19	Arthralgia	0.18	0.18
Headache	0.18	0.19	Nausea	0.20	0.18
Pain Limb	0.16	0.18	Pain Abdominal	0.21	0.16

Conclusions:

The results of the two phase III studies show that the change in hemoglobin levels from the baseline was within the prospectively set limit. In this regard, both studies have met the efficacy criterion. The results also hold within subsets of sex and age groups. These results are slightly different then provided by the sponsor. For the pivotal phase III study 980117, our results are based on the three datasets provided by the sponsor. All three datasets yielded similar results of non-inferiority. Difference between our results and the results of the sponsor are slightly different due to different number of patients used. The dataset called 'observed' is not exactly same as what the sponsor calls 'per protocol', the primary analysis. The data set 'observed' has slightly more patients then 'per protocol' analysis. The sponsor has also provided several additional analyses based on modified ITT data set and several variations of it and with and without adjustment with covariables. All these data resulted in similar conclusions. Therefore, slightly different

results obtained by the reviewer were not considered of much concern. All analyses lead to the same conclusion that the NESP is not inferior to EPO.

Both studies also showed slightly larger within patient variability in NESP group. For the US/Canada study, variability was close to significance level. For the European study, the results were not statistically significant. The safety profiles of the most common adverse experiences were also found similar in both treatment groups.